R E P O R T R E S U M E S

ED 016 957 48 AL 000 927

THE PATTERN OF AIR FLOW OUT OF THE MOUTH DURING SPEECH.

BY- LANE, H. AND OTHERS
MICHIGAN UNIV., ANN ARBOR,CTR.FOR RES.LANG.AND BEH

REPORT NUMBER BR-6-1784 PUB DATE SEP 67
CONTRACT OEC-3-6-061784-0508
EDRS PRICE MF-$0.25 HC-$0.64 14P.

DESCRIPTORS- *ARTICULATION (SPEECH), BEHAVIORAL SCIENCE RESEARCH, RESEARCH METHODOLOGY, COMPUTER PROGRAMS, *ACOUSTIC PHONETICS, RESEARCH PROBLEMS, AUTOMATIC SPEECH RECOGNITION, KYMOGRAPHIC RECORDINGS, AIR FLOW PATTERNS,

SINCE THE 19TH CENTURY, KYMOGRAPHIC RECORDING OF TOTAL AIR FLOW OUT OF THE MOUTH HAS BEEN USED TO DIAGNOSE THE VARYING DURATIONS AND DEGREES OF CONSTRICTIONS OF THE VOCAL TRACT DURING SPEECH. THE PRESENT PROJECT ATTEMPTS TO INTRODUCE A SECOND DIMENSION TO RECORDINGS OF AIR FLOW OUT OF THE MOUTH--NAMELY, CROSS-SECTIONAL AREA OF FLOW--ON THE HYPOTHESIS THAT THIS WILL REFLECT CHANGES IN THE LOCATION AND CROSS-SECTIONAL SHAPE OF CONSTRICTIONS OF THE VOCAL TRACT. CONSEQUENTLY, THE FINDINGS ARE PERTINENT TO AUTOMATIC SPEECH RECOGNITION AND ALLIED OBJECTIVES. THE PROCEDURE INCLUDED THE DEVELOPMENT OF A MATRIX OF 64 HOT-WIRE ANEMOMETERS AND ASSOCIATED CIRCUITRY, WHERE OUTPUT IS SAMPLED AND DIGITIZED FOR COMPUTER PROCESSING. COMPUTER PROGRAMS PERMIT STORAGE, AVERAGING, TRANSPORTATION, NORMALIZATION AND MATCHING OF AIR FLOW PATTERNS, AS WELL AS SUSTAINED CRT DISPLAY AND NUMERICAL READOUT. THE PATTERNS ASSOCIATED WITH VARIOUS SOUNDS AND THEN WITHIN- AND BETWEEN-SPEAKER VARIANCE ARE REPORTED. THIS PAPER WAS PRESENTED TO THE ACOUSTICAL SOCIETY, LOS ANGELES, CALIFORNIA, NOVEMBER 4, 1966, AND ALSO APPEARS IN "STUDIES IN LANGUAGE AND LANGUAGE BEHAVIOR, PROGRESS REPORT V," SEPTEMBER 1, 1967. (AUTHOR/AMM)

The Pattern of Air Flow Out of the Mouth During Speech[1, 2]

H. Lane, J. C. Catford, R. Oster,
F. E. O'Donnell, and T. Rand
Center for Research on Language and Language Behavior
University of Michigan

Since the 19th century, kymographic recording of total air flow out of the mouth has been used to diagnose the varying durations and degrees of constrictions of the vocal tract during speech. The present project attempts to introduce a second dimension to recordings of air flow out of the mouth--namely, cross-sectional area of flow--on the hypothesis that this will reflect changes in the location and cross-sectional shape of constrictions of the vocal tract. Consequently, the findings are pertinent to automatic speech recognition and allied objectives. The procedure included the development of a matrix of 64 hot-wire anemometers and associated circuitry, where output is sampled and digitized for computer processing. Computer programs permit storage, averaging, transportation, normalization and matching of air flow patterns, as well as sustained CRT display and numerical readout. The patterns associated with various sounds and then within- and between-speaker variance are reported.

Since man often understands what is said to him, and this ability is useful, it has also seemed useful to many investigators to try to develop a machine that would understand what was said to it. The question that is typically so easy for the listener to answer--namely, what was said--has proven difficult for a machine to answer. Past investigators have sought to wring out of the acoustic waveform a frank admission of its own identity. We have filtered, modulated, rectified, segmented, peak clipped, autocorrelated, stretched, compressed, nonlinearly amplified, but through all, the waveform kept its own counsel. The more the speech waveform was uncommunicative, the more we have maimed, mangled and mutilated it, apparently in the hope that the stimulus for speech recognition would become recognizable if only it were distorted enough--and in the right ways. Because our machines still find it so difficult to do what our listener finds so easy, some wag ventured the thought that the listener may bring to bear on this task his years of training as a listener; another pointed out that the listener has considerable experience as a speaker; and a third hinted darkly that the listener's genetic and physiological endowment may be critically involved.

Since a complete account of language acquisition and its biological bases had not yet appeared at the time we began our research, we pursued a different

route out of the box. We reasoned in this way: since we do not know enough about how a listener processes speech to simulate this in a device, perhaps we have been wrong in assuming that just because the listener processes the acoustic waveform for speech recognition, a recognition device must process the acoustic waveform, too. Instead, there may be other correlates of the vocal response (which must therefore be correlated with the acoustic waveform) that may be more recognizable for the machine, even if they are unrecognizable for the listener. One such correlate, which has received almost no public attention in this regard, is the pattern of air flow out of the mouth during speech.

Since the latter half of the 19th century the kymographic recording of air flow out of the mouth has been a basic tool of phonetic research. Originally using rubber diaphragms and tracings on smoked paper wrapped around a clockwork-driven drum, kymography now uses electronic equipment; but the objects and results are essentially the same, namely to record variations in air flow which are diagnostic of phonatory and articulatory events in the vocal tract.

The linear trace of the kymograph, however, gives, in effect, information about only one dimension of vocal activity, namely the varying durations and degrees of constrictions of the tract, as indicated by the variations through time of the volume-velocity of air flow out of (or into) the mouth and nose.

The present research project is an attempt to introduce a second dimension into recordings of air flow out of the mouth, namely, to add at least some indications of the location and cross-sectional shape of constrictions of the vocal tract. We have a report of current research in progress: an account of some of our methodological mistakes and advances, some initial findings, and an indication of where we are headed next.

The hypothesis is that articulatory constrictions at different locations and of different cross-sectional shape and area should in many cases produce differences in the location and cross-sectional area of the total mass of air flowing out of the mouth. Thus, it might be predicted that air flowing through a lateral channel (as for /l/) would show a more laterally-located outflow than air flowing through a central channel (as for /z/), that air flowing through narrowly-rounded lips, as for /u/. would show a more concentrated and centrally located outflow than air flowing through a vertically wide opening as for /a/, and so on. It was further hoped that finer discriminations might be obtained. For instance, there might be characteristic differences in the cross-sectional

flow pattern for such pairs of sounds as /s/ and /s̬/; the latter, perhaps, having the flow directed somewhat more downwards than the former.

It was realized from the start that there might be at least three possible sources of problems. First, the Ss were to produce isolated sounds representing as far as possible certain English phonemes. However, phonemes cannot be defined absolutely but only relative to other phonemes; the articulators are allowed some degree of freedom in their placement. Thus, an English /t/ is recognized as such whether the tongue is touching the teeth or the back part of the alveolar ridge. In another language these two positions might define two rather than just one phoneme. Thus, when a S attempts to say a sound twice, representing the same phoneme, a slight change in articulation might change the air flow pattern but not the phoneme. This factor would be even more apparent between Ss because each person tends to develop his own characteristic way of producing each phoneme. Second, the size and shape of the vocal tract varies from person to person. Thus, because of physiological limitations, even approximately identical articulatory positions in different people may not produce identical air flow patterns. Third, the initial air flow, usually from the lungs, may vary without changing the quality of the sound. This variation could also contribute to a variation in air flow pattern.

Despite all these possible complications, it must be remembered that for the present purposes it is not necessary that a given phoneme result in exactly the same air flow pattern each time. It is only necessary that the pattern be distinguishable from the patterns formed by all other phonemes in the set employed.

Initial experiments used a moveable hot-wire anemometer suspended in front of the mouth from eyeglasses. This first foray was inconclusive, not only because of the problems mentioned above, but also for two other reasons.

Insert Figure 1 about here

First, the necessity of placing the anemometer probe successively in four different positions meant that for each sound studied we obtained recordings of four different utterances. It was, therefore, difficult to draw conclusions from these data about the air-velocity distribution over the four probe-locations for any one utterance of a sound.

Secondly, in using only four probe-locations it was impossible to tell if these locations were optimal. In any given case we might, in fact, have been

missing the locus or loci of maximal flow velocity, or otherwise failing to get a reliable picture of the cross-sectional pattern of the flow.

It became necessary, therefore, to find a way of obtaining a more general visualization of the location of the whole mass of air flowing out of the mouth.

Method

Several methods of flow-visualization were considered. These included methods using infra-red heat and moisture-sensitive chemicals and also the Schlieren technique. The latter method uses the deflection of parallel light-rays by a density gradient, provided in this case by articulating after inhaling helium.

Insert Figures 2 and 3 about here

Next, a method using liquid nitrogen was explored. A copper screen was dipped in liquid nitrogen and then placed on a rigid frame. Because of the rapid evaporation of the liquid nitrogen, the screen became quite cold and moisture from the air formed a frost layer on it. The gradual return of the screen's temperature toward that of the room was monitored with a circuit including a constantan wire soldered to the center of the copper screen at one end and to a copper lead, submerged in ice-water, at the other. A VTVM measured the voltage between the lead and the screen, which was the sum of the emf generated at the two thermocouple junctions, and was proportional to the temperature of the screen referred to that of the ice water. Through experimentation, a temperature was selected at which the screen appeared to be most sensitive to the air flow from the mouth (-5°C). Although this threshold was low, its effect on the measured pattern can probably not be neglected.

The S approached the screen just before it reached threshold temperature. He placed his nose against a wire projection, which positioned his lips approximately 1.5 cm from the screen, and then he produced the given phoneme. The screen began to defrost from the warmth of the air flowing from his mouth. By use of a voice-operated relay, a photograph of the pattern on the screen was taken exactly 1.5 sec. after the S began to phonate. This time lapse allowed

---

Insert Figure 4 about here

---

the pattern to develop but was not so long that defrosting from other factors had much chance to alter the pattern. Each observed pattern of air flow is, therefore, time integrated; it is not the pattern at any given moment during production of the sound. The utterance was simultaneously recorded on a tape recorder for later measurement of relative sound levels.

Seven sounds, representing approximately isolated utterances of the English phonemes /a/, /u/, /l/, /s/, /$\overset{v}{s}$/, /z/, and /$\overset{v}{z}$/ were used as samples. In a given session a S spoke each of the seven sounds twice in succession; he tried to duplicate each sound as closely as possible in articulatory position and effort. Ss 1 and 2 each served in two sessions a week apart, while S 3 served only once.

Results and Discussion

Figures 5, 6, and 7 present tracings of the photographs of the air flow patterns. The + represents that point in the S's mouth where the lower edges of the two upper front teeth come together. The patterns marked by solid lines

---

Insert Figures 5, 6, and 7 about here

---

and dashed lines were obtained in the first session; those marked by dash-dot-dash and dash-plus-dash were recorded in the second.

Examination of the tracings reveals that the pattern of air flow is quite consistent over utterances of the same phoneme by one S. In this respect, the vowels are the most consistent. Their patterns are very much like those that might be predicted from general phonetic knowledge. The patterns for /a/ followed very closely the outline of the shape of the mouth. The absence of a record here for S 2 is understandable because the velocity of air flow from the mouth during this sound is low. However, it should be noted that no systematic relation was observed between the area of the air flow patterns and the sound pressure level of their correlated sounds. The patterns for/u/ for all Ss were small and centered about the mouth. This type of pattern is consistent with the rapid air flow, the rounding of the lips, and the small oral opening that occur during this sound.

The patterns for /l/ are particularly interesting. Ss 2 and 3 show clearly unilateral patterns--with the difference that 2's /l/ appears to be left-sided, 3's /l/ right-sided. S 1's /l/ is apparently bilateral, and hence not so distinct from /z/ and /$\overset{v}{z}$/ as that of 2 and 3.

The voiced fricatives seem to be the next most consistent within a S. The voiceless fricatives show the widest variation, which may be partly the result of their very turbulent air flow.

There seems to be enough variation between patterns of different sounds for a given S for purposes of differentiation. The variations between sounds seem to be greater than the variations within a sound.

In contrast to the intra-subject consistency, there is limited inter-subject consistency for a given phoneme, especially in the case of consonants. We concluded that an efficient pattern-matching method for the differentiation of phonemes is quite possible, based on their correlated patterns of air flow, for a given S but that, because of the extensive inter-subject variation, the standard patterns would have to be modified for each new S.

In order to proceed with pattern matching of air flow patterns, and to explore whether there might be some transformations of these patterns which would make the data from various Ss more congruent, we constructed a matrix of 16 hot-wire anemometers, positioned in front of the mouth. Through the use of a multiplexor and A-D converter, a digital measure of the velocity of the air passing by each anemometer during articulation is stored in our PDP-4 computer. The whole matrix is sampled approximately 120 times per sec. Since there is storage available for approximately 112 matrices, at any given time, approximately 2 sec. of air flow are stored.

The computer program for this conversion can also perform the following functions. It can display the matrix on an oscilloscope in real time with intensity as a measure of velocity; so that the pattern of air flow is readily seen. It can hold the current matrix on the oscilloscope screen. The program can yield all matrices in storage. It can also sum the air flow for each matrix and print these sums or display them, through the D/A converter on a graphic recorder. This graphic record is the typical flow function of time usually obtained with a face mask that collects the flow and a single transducer that detects its velocity. Finally, the program can yield an average matrix for occurrences of isolated sounds by finding the matrix with maximum air flow and averaging it with the two preceding and following matrices.

Programs are currently being written for converting these average matrices of isolated sounds into binary matrices through the use of a flow threshold (cut off value). All the average binary matrices for a given sound will then

be averaged to produce a characteristic binary matrix for each sound. Various methods of pattern-matching will then be used to match matrices of unknown sounds with those of known sounds. The basic pattern-matching procedure will probably be a 1-1 comparison of elements of the matrices--at least, initially.

Footnotes

[1]This paper was presented to the Acoustical Society, Los Angeles, California, November 4, 1966.

[2]The research reported herein was performed pursuant to Contract OEC-3-6-061784-0508 with the U. S. Department of Health, Education, and Welfare, Office of Education, under the provisions of P. L. 83-531, Cooperative Research, and the provisions of Title VI, P. L. 85-864, as amended. This research report is one of several which have been submitted to the Office of Education as *Studies in language and language behavior*, Progress Report V, September 1, 1967.

Figure Captions

Fig. 1. Moveable hot-wire anemometer probe and reference probe (R) suspended in front of the mouth.

Figs. 2-3. Flow-visualization by the Schlieren technique: the white cloud indicates the flow of helium out of the mouth during the production of /s/ (2) and /ʃ/ (3).

Fig. 4. A photograph illustrating how patterns of air flow were detected using a frosted screen.

Figs. 5-7. Tracings of air flow patterns taken from photographs like that in Fig. 4.

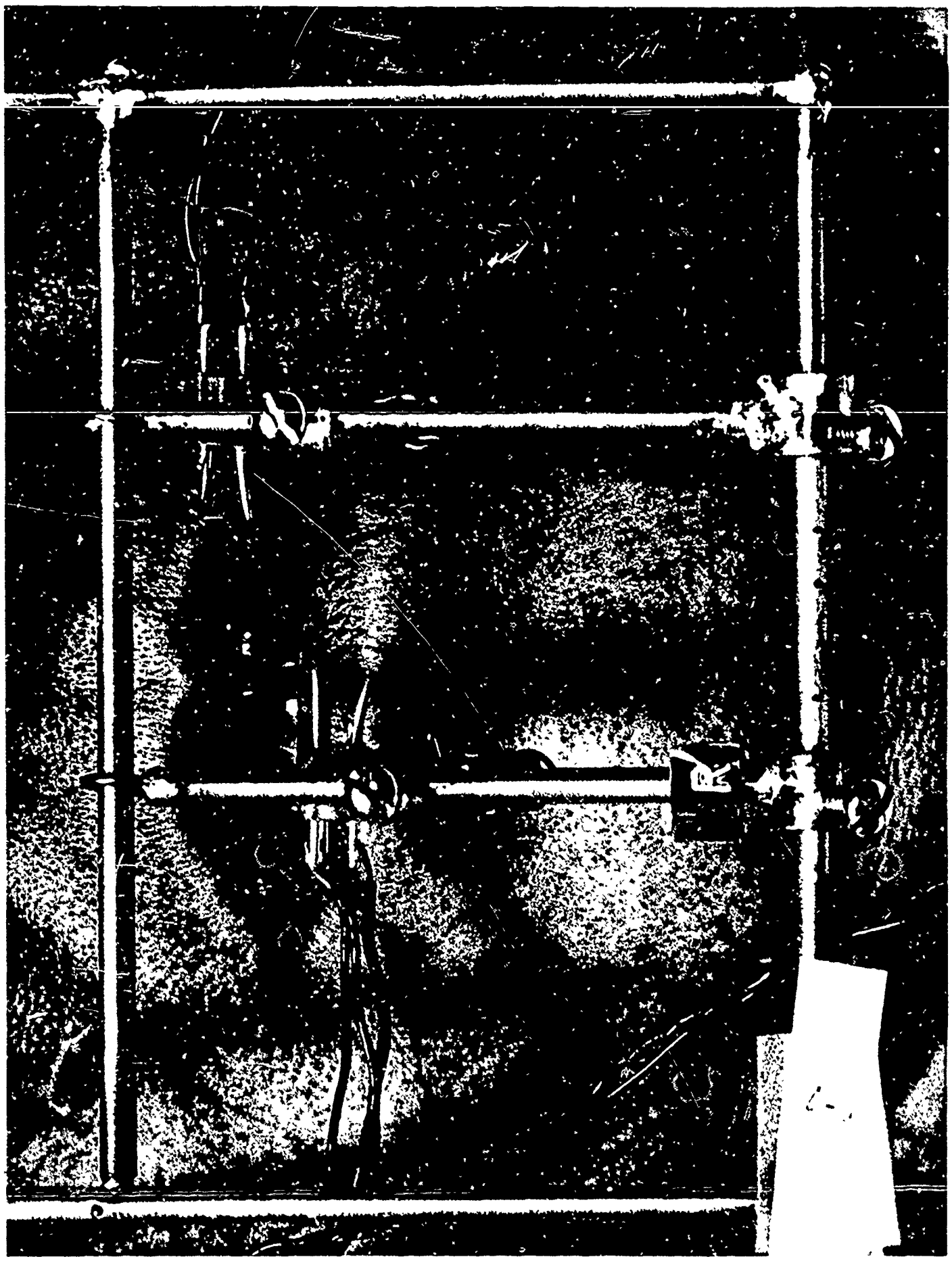

Fig. 1

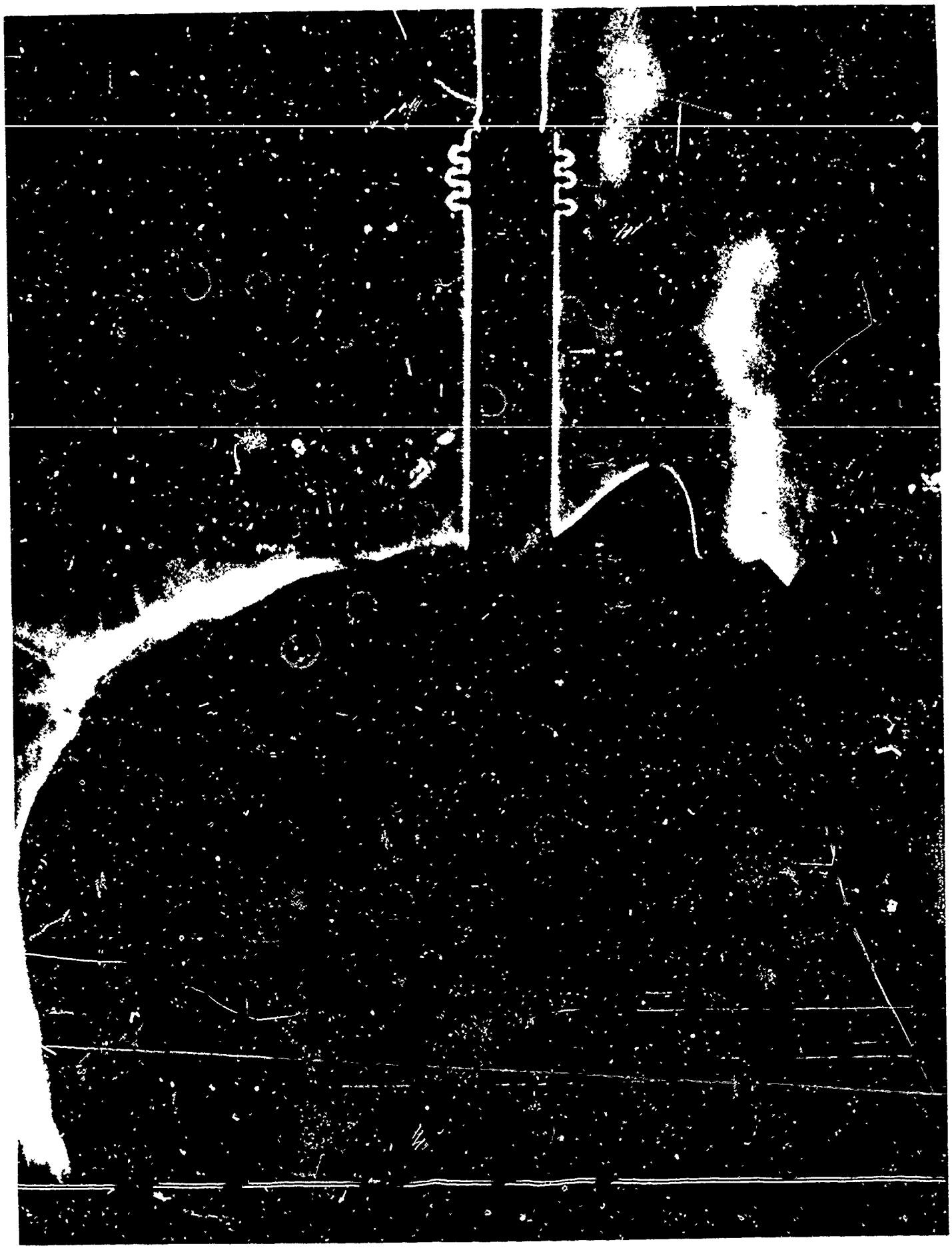

Fig. 2

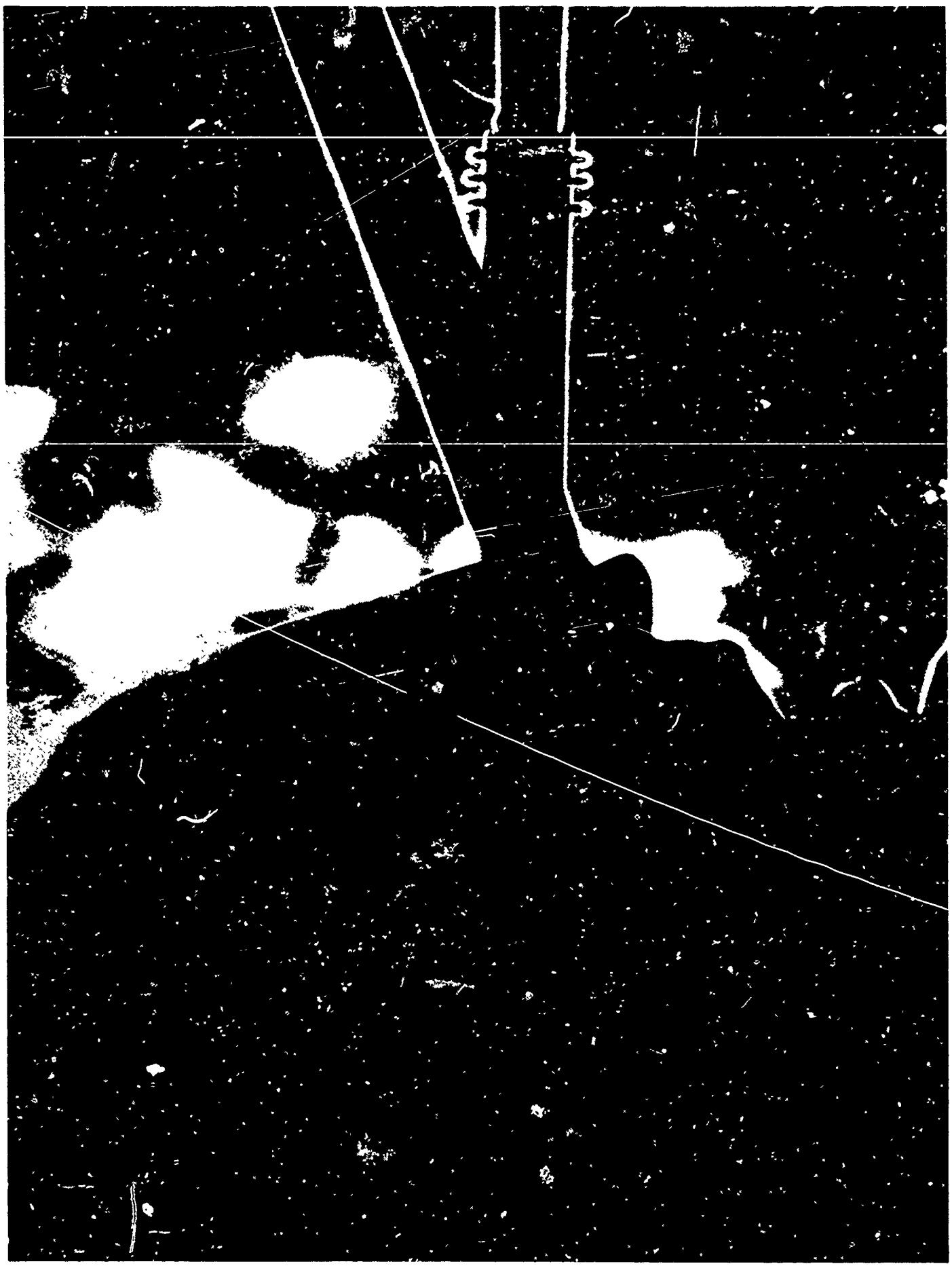

Fig. 3

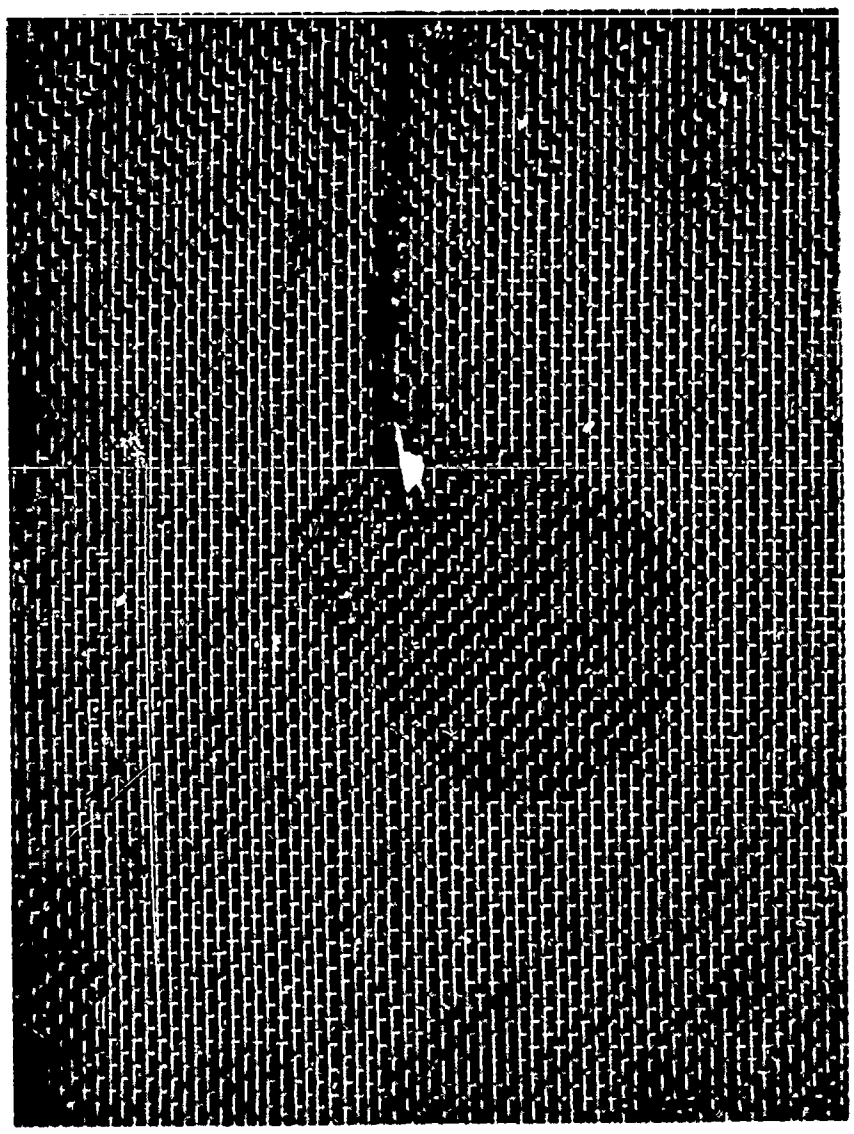

Fig. 4

PHONEME /a/
PHONEME /u/
PHONEME /l/
Subject One
(No patterns)
Subject Two
Subject Three

Fig. 5

Fig. 6

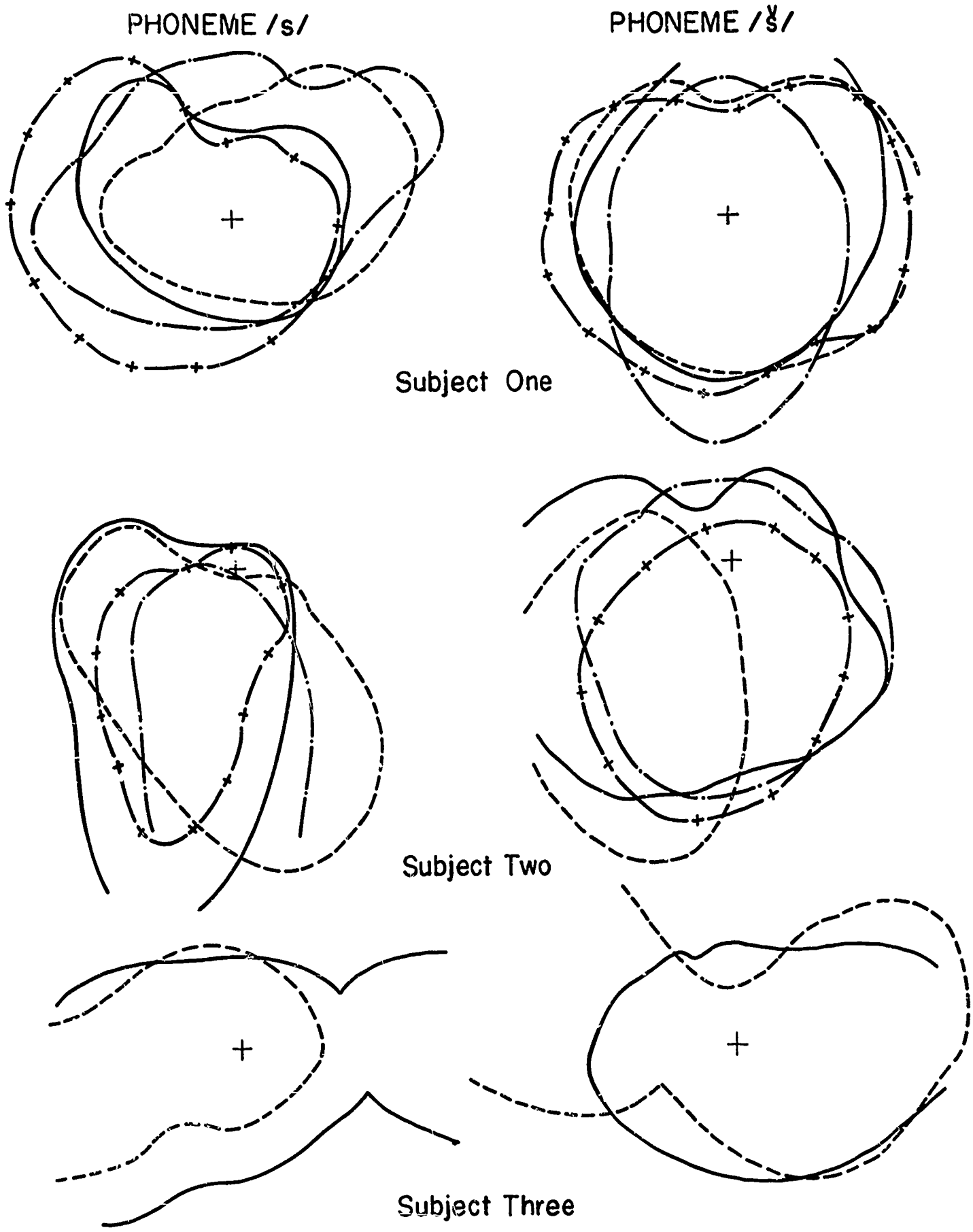

Fig. 7